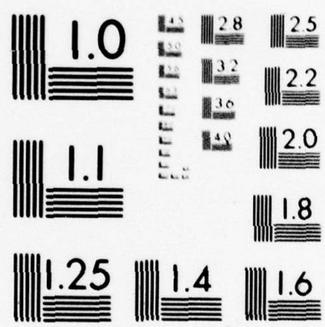
AD-A072 991 SOUTHERN METHODIST UNIV DALLAS TEX DEPT OF STATISTICS F/G 12/1
A CONDITIONAL PROPERTY OF ADAPTIVE ESTIMATORS.(U)
JUL 79 W C PARR N00014-75-C-0439
UNCLASSIFIED TR-133 NL

1 OF 1

AD
A072991

END
DATE
FILMED
9-79
DDC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL

12

SOUTHERN METHODIST UNIVERSITY

See 1473 in back

DEPARTMENT OF STATISTICS

DALLAS, TEXAS 75275

12

A CONDITIONAL PROPERTY OF ADAPTIVE ESTIMATORS

by

William C. Parr
Institute of Statistics, Texas A&M
University, College Station, Texas

Technical Report No. 133
Department of Statistics ONR Contract

July 25, 1979

Research Sponsored by the Office of Naval Research
Contract N00014-75-C-0439
Project NR 042-280

DIVISION OF MATHEMATICAL SCIENCES
Department of Statistics
Southern Methodist University
Dallas, Texas 75275

A CONDITIONAL PROPERTY OF ADAPTIVE ESTIMATORS

By William C. Parr

Institute of Statistics, Texas A&M University, College Station, Texas

SUMMARY

In adaptive estimation, it is often considered that an estimator has made a mistake if the component estimator chosen for use is not the most efficient for the distribution sampled. Theoretical and simulation results point to a fallacy in this line of thought. The Monte Carlo study involves extension of the Princeton Swindle to distributions conditional on a location- and scale-free statistic, and to the uniform. The results give a partial explanation for the sometimes surprising robustness of adaptive L-estimators.

Keywords:- Adaptive estimation; conditional reference sets; Monte Carlo swindles; robustness

1. INTRODUCTION

There has been a great deal of recent interest in the use of adaptive estimators to achieve the goals of robust estimation. Hogg (1974) gives a broad perspective on the state of the art from one of the field's pioneers. Adaptive estimators may be roughly characterized as follows.

A "preliminary" statistic (Hogg's Q as defined below, the sample kurtosis, or perhaps some combination of skewness and tailweight measures) is calculated from the data, and used to select an estimator T from a class $\{T_\lambda, \lambda \in \Lambda\}$. Typically, the preliminary statistic is a sample measure of a population characteristic known to affect drastically the relative behavior of the T_λ.

The motivation for our study of the behavior of estimators conditional on a preliminary statistic is best understood by means of an example. Let the adaptive estimator T (for the point of symmetry of a symmetric population) be defined by

$$T = \begin{cases} T(.10) & \text{if } Q < 2.0 \\ T(.25) & \text{if } Q \geq 2.0 \end{cases}$$

where

$$T(\alpha) = \begin{cases} \text{median} & \alpha = .5 \\ \dfrac{1}{n-2[n\alpha]} \displaystyle\sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{(i)} & 0 \leq \alpha < .5 \\ \dfrac{1}{2\{[n(\alpha+.5)]+1\}} \displaystyle\sum_{i=1}^{[n(\alpha+.5)]+1} (X_{(i)} + X_{(n-i+1)}) & -.5 < \alpha < 0 \\ \text{midrange} & \alpha = -.5 , \end{cases}$$

$X_{(i)}$ $(i = 1, \ldots, n)$ are the sample order statistics, and

$$Q = \frac{\bar{U}(.05) - \bar{L}(.05)}{\bar{U}(.5) - \bar{L}(.5)} ,$$

$\bar{U}(\alpha)$ and $\bar{L}(\alpha)$ being the average of the largest and smallest $[n\alpha]$ order statistics respectively. Note that, for $\alpha < 0$, $T(\alpha) = \frac{1}{2}\{\bar{U}(|\alpha|) + \bar{L}(|\alpha|)\}$. Q is thus a location-and scale-free tailweight measure. (Population values are 1.95, 2.58, and 3.30 for the uniform, normal, and double exponential, respectively.)

The usual presumption has been that if one is sampling from an extremely heavy-tailed distribution and observes a sample with $Q < 2.0$, hence using $T = T(.10)$, a 10% trimmed mean instead of a 25% trim, an error has been made. The following work demonstrates, however, that it is quite possible that although

$$\mathrm{Var}_{\mathrm{Cauchy}}(T(.10)) > \mathrm{Var}_{\mathrm{Cauchy}}(T(.25)) \,,$$

and also, in fact,

$$\mathrm{Var}_{\mathrm{Cauchy}}(T(.10)\,|\,Q < 2.0) > \mathrm{Var}_{\mathrm{Cauchy}}(T(.25)\,|\,Q < 2.0) \,,$$

$T(.25)$ may be much less inefficient with respect to $T(.10)$ when sampling from the conditional distribution, given $Q < 2.0$. This would agree with the view that samples from the Cauchy with $Q < 2.0$ are more uniform- or normal-like than their unconditional counterparts.

2. SOME THEORETICAL INSIGHTS

In this section we consider the case where K is a location- and scale-free statistic, T is an unbiased location estimator (i.e. $T[a\underline{X} + b\underline{1}] = aT[\underline{X}] + b$ for $a \neq 0$), and the sampling is from a location family, i.e. $f_\theta(x) = f(x - \theta)$ for all $x \in R$. If a complete and sufficient statistic for θ exists, we let T_{cs} denote the unbiased estimator of θ which is a function of it.

Then,

$$\begin{aligned} Var(T|K) &= Var(T - T_{cs} + T_{cs}|K) \\ &= Var(T - T_{cs}|K) + Var(T_{cs}|K) \\ &= E\{(T - T_{cs})^2|K\} + Var(T_{cs}) \end{aligned}$$

by Basu's theorem, since the distribution of $T - T_{cs}$ is parameter free. (Note that Basu's theorem works for the conditional distributions because T_{cs} and K are statistically independent, as are T_{cs} and $T - T_{cs}$.)

Thus, we see that

(i) If $T \equiv T_{cs}$, i.e. if our estimator is a function of a complete and sufficient statistic, then the conditional variance (in fact the entire distribution) of T is the same as the unconditional.

(ii) A reasonable measure of the extent the conditional variance of T depends upon K would be

$$Var[E\{(T - T_{cs})^2|K\}] = \tau$$

and τ would be made large by a dependence between K and $|T - T_{cs}|$, together with the possibility of large values of $|T - T_{cs}|$.

Based upon these two considerations, we expect the conditional variances to vary from the unconditional primarily for estimators T which are of low efficiency and for conditioning variables K which truly reflect characteristics which are related to the efficiency of T.

3. SIMULATION METHODS AND RESULTS

Several points regarding the Monte Carlo methods used in this study are of interest. It is desired to estimate the variances of $T(\alpha)$ for several values of α, at the normal, $.75N(0,1) + .25N(0,9)$, and uniform distributions, conditional on observing Q in specified ranges, for samples of size 20. (Note $.75N(0,1) + .25N(0,9)$ is not a true distribution, but a pseudo-sample consisting of 15 observations from $N(0,1)$ and 5 from $N(0,9)$.)

The first point is that, for those distributions where the Princeton Swindle is applicable, the extension to those distributions conditional on a location-and scale-free statistic is immediate. If $X_i = Z_i/Y_i (i = 1, ..., n)$ where the Z_i are iid $N(0,1)$ and the Y_i are such that $P[Y_i = 0] = 0$ $(i = 1, ..., n)$, then $a = \Sigma Z_i Y_i / \Sigma Y_i^2$ and $b = [\{\Sigma Z_i^2 - (\Sigma Z_i Y_i)^2/\Sigma Y_i^2\}/(n-1)]^{\frac{1}{2}}$ are, conditional upon $\underline{Y}$, statistically independent of $\underline{c} = (\underline{x} - \underline{1}a)/b$. Therefore, if K is a location and scale free statistic, it is a function of $\underline{c}$ and thus independent of a and b, which yields

$$\mathrm{Var}(T|K) = E[a^2] + \mathrm{Var}(T(\underline{c})|K)$$

by the usual manipulations. Thus, we will estimate the conditional variance $\mathrm{Var}(T|K)$ by

$$\hat{\mathrm{Var}}(T|K) = E[a^2] + \frac{1}{M}\sum_{i=1}^{M} T^2(\underline{c}_i) ,$$

where we have simulated the unconditional distribution N times, of which M satisfied the conditioning criterion. Arguments similar to those of Gross (1973) and Simon (1976) demonstrate that this swindle can only serve to reduce Monte Carlo variability. For distributions representable as N(0,1)/Independent, we will thus generate from the unconditional distribution 10000 samples, estimating the various conditional quantities desired based upon those of the 10000 samples meeting the criterion.

The Princeton Swindle has not, however, been extended to short-tailed distributions. Nevertheless, an analogous variance reduction technique is available for any location-scale family admitting a two-dimensional complete and sufficient statistic (a,b), where a is a location statistic $\{a(c_1\underline{x} + c_2\underline{1}) = c_1 a(\underline{x}) + c_2,\ c_1 \neq 0,\ -\infty < c_2 < \infty\}$, and b is a scale statistic $\{b(c_1\underline{x} + c_2\underline{1}) = |c_1| b(\underline{x}),$ for $c_1 \neq 0$, all $-\infty < c_2 < \infty\}$. Under these conditions, and with K as above,

$$\mathrm{Var}(T|K) = E[a^2] + E[b^2]\, E[T^2(\underline{c})|K] ,$$

which permits essentially the same procedure as when the Princeton Swindle applied. Thus, in the case of sampling from a uniform distribution on the interval (-1,1),

$$\mathrm{Var}(T|K) = \frac{2}{(n+1)(n+2)} + \frac{n(n-1)}{(n+1)(n+2)}\, E[T^2(\underline{c})|K] \quad .$$

It should be noted that the savings due to the swindle in this case will be $O(\frac{1}{n^2})$ (order of the theoretically calculated term), an order in n higher than $E[T^2(\underline{c})|K]$ for most estimators. However, for near-efficient estimators, especially those with $E[T^2(\underline{c})|K] = O(\frac{1}{n^2})$, the gains will be especially valuable for small n. In independent work, Beal (1974) has also extended the Princeton Swindle to the case of a two-dimensional complete and sufficient statistic (a,b) with a a location statistic and b a scale statistic, but not to the case of conditioning upon a location- and scale-free statistic.

Normal pseudorandom variates were generated by means of the polar method, with a multiplicative congruential uniform generator used to obtain pseudorandom uniforms. Tables 1, 2, and 3 give estimated variances for $n = 20$ of $T(\alpha)$ for $\alpha = -.5, -.4, -.25, 0, .05, .10, .25$, and $.50$, conditional on $Q < 2.0$, $2.0 \leq Q \leq 2.6$, $2.6 < Q$, and unconditionally for the $N(0,1)$, $.75N(0,1) + .25N(0,9)$, and $U(-1,1)$ respectively. These ranges correspond roughly to the categories for Q used in T_1 of Hogg(1974), the first, second and third categories containing the population values for the uniform (1.95), normal (2.58), and double exponential (3.30) respectively. The entries are $n \times$ (estimated variance), followed in parentheses by an estimate of the standard error of these simulated values. The bottom row in each table gives the number of the 10000 samples falling into the conditional or unconditional category.

(Tables 1, 2, and 3 about here)

Several points emerge as worthy of note from an examination of these tables. First is that our suggestions from section 2 hold true. The highly efficient estimators (trims with small α for $N(0,1)$, α near .25 for $.75N(0,1) + .25N(0,9)$, and α near $-.5$ for $U(-1,1)$) exhibit only small or no deviations of their conditional variances from their unconditional variances. However, highly inefficient estimators (α near $-.5$ for $N(0,1)$ or $.75N(0,1) + .25\ N(0,9)$, and α near .5 for $U(-1,1)$) deviate greatly in this regard. For instance, the midrange ($\alpha = -.5$), while having an unconditional variance over 19 times that of $T(.25)$ for $.75N(0,1) + .25N(0,9)$, has the same ratio approximately equal to 2.7 conditional on $Q < 2.0$. Thus, usage of $T(-.5)$ wherever $Q < 2.0$ would be not nearly as serious an error in this sampling situation as its unconditional behavior would suggest. Similar observations hold for other comparisons.

A further use of these results (and, more generally speaking, this sampling methodology) would be to assess the behavior of possible adaptive estimators without extensive costly Monte Carlo studies. For instance, for the estimator

$$T = \begin{cases} T(-.5) & Q < 2.0 \\ T(0) & 2.0 \leq Q \leq 2.6 \\ T(.5) & 2.6 < Q \end{cases}$$

(admittedly not a terribly informed choice) we estimate the variance at N(0,1) (all variances still multiplied by 20, the sample size) to be

$$\hat{\text{Var}}(T) = \frac{935}{10000}(1.578) + \frac{6648}{10000}(1.000) + \frac{2417}{10000}(1.372) = 1.144\,.$$

Similarly, at .75N(0,1) + .25N(0,9) T would have estimated variance 2.283 and at U(-1,1) it would have estimated variance .211. An analysis incorrectly ignoring the conditional behavior (taking the same linear combination as above but using the unconditional variances) would have estimated these three variances as 1.287, 2.732, and .176. Thus, we see that for the first two distributions, an assessment based upon unconditional behavior drastically underrates the adaptive estimator. The situation is, however, reversed for the (admittedly extreme) uniform distribution. Thus we see that naive assessment of adaptive estimators may well seriously misestimate (generally overestimate) their variances.

4. SUMMARY AND CONCLUSIONS

Theoretical and Monte Carlo results are given to suggest that the reason for the surprising robustness of adaptive estimators may lie in the unsuspectedly good conditional behavior of their components. To

facilitate efficient simulation, variance reduction methods were extended to the relevant conditional situations. The results provide a method for assessing more accurately and efficiently the behavior of prospective robust adaptive estimators.

5. ACKNOWLEDGMENT

Portions of this research were supported by ONR Contract N00014-75-C-0439.

REFERENCES

Beal, S. (1974) Adaptive M Estimation With Independent Nonidentically Distributed Data. Unpublished dissertation, UCLA.

Gross, A. M. (1973) A Monte Carlo Swindle for Estimators of Location. Applied Statistics, 22, 347-353.

Hogg, R. V. (1974) Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory. J. Amer. Statist. Assoc., 69, 909-927.

Simon, G. (1976) Computer Simulation Swindles, with Applications to Estimates of Location and Dispersion. Appl. Statist., 25, 266-274.

TABLE 1

Estimated Conditional Variances of T(α) when Sampling from N(0,1)

	α	Q < 2.0	2.0 ≤ Q ≤ 2.6	2.6 < Q	Unconditional
Midrange	-.5	1.578 (.025)	2.408 (.024)	4.096 (.067)	2.871 (.027)
	-.4	1.411 (.019)	1.790 (.014)	2.167 (.027)	1.876 (.012)
	-.25	1.158 (.007)	1.190 (.003)	1.180 (.004)	1.184 (.003)
Mean	0.0	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)
	.05	1.007 (.000)	1.017 (.000)	1.038 (.000)	1.023 (.000)
	.10	1.026 (.001)	1.049 (.001)	1.073 (.002)	1.055 (.001)
	.25	1.158 (.007)	1.190 (.003)	1.180 (.004)	1.184 (.003)
Median	.50	1.621 (.027)	1.490 (.009)	1.372 (.009)	1.464 (.006)
# Repetitions		935	6648	2417	10,000

TABLE 2

Estimated Conditional Variances of T(α) when Sampling from .75N(0,1) + .25N(0,9)

	α	Q < 2.0	2.0 ≤ Q ≤ 2.6	2.6 < Q	Unconditional
Midrange	-.5	5.045 (.509)	13.474 (.387)	39.934 (.529)	34.610 (.443)
	-.4	4.812 (.501)	10.999 (.327)	20.929 (.277)	18.856 (.234)
	-.25	3.530 (.350)	5.592 (.155)	6.616 (.080)	6.382 (.070)
Mean	0.0	2.334 (.139)	2.992 (.056)	3.056 (.026)	3.034 (.023)
	.05	2.192 (.115)	2.550 (.041)	2.253 (.015)	2.307 (.014)
	.10	2.059 (.091)	2.221 (.030)	1.883 (.010)	1.948 (.010)
	.25	1.879 (.063)	1.952 (.020)	1.766 (.007)	1.803 (.007)
Median	.50	2.122 (.094)	2.286 (.031)	2.070 (.012)	2.111 (.011)
# Repetitions		138	1864	7998	10,000

TABLE 3

Estimated Conditional Variances of T(α) when Sampling from U(-1,1)

	α	$Q < 2.0$	$2.0 \le Q \le 2.6$	$2.6 < Q$	Unconditional
Midrange	-.5	.087 (.000)	.087 (.000)	.087 (.000)	.087 (.000)
	-.4	.099 (.000)	.122 (.001)	.241 (.016)	.108 (.001)
	-.25	.157 (.001)	.253 (.004)	.482 (.044)	.192 (.002)
Mean	0.0	.282 (.003)	.434 (.008)	.676 (.071)	.336 (.003)
	.05	.328 (.004)	.516 (.010)	.815 (.087)	.395 (.004)
	.10	.378 (.005)	.591 (.012)	.866 (.094)	.452 (.005)
	.25	.553 (.007)	.767 (.016)	.979 (.109)	.627 (.007)
Median	.50	.844 (.012)	.922 (.020)	1.086 (.132)	.872 (.011)
# Repetitions		6641	3250	109	10,000

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 133
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): (6) A CONDITIONAL PROPERTY OF ADAPTIVE ESTIMATORS
5. TYPE OF REPORT & PERIOD COVERED: (9) Technical Report
6. PERFORMING ORG. REPORT NUMBER: 133
7. AUTHOR(s): (10) William C. Parr
8. CONTRACT OR GRANT NUMBER(s): (15) N00014-75-C-0439
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Southern Methodist University, Dallas, TX 75275
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 042 280
11. CONTROLLING OFFICE NAME AND ADDRESS: OFFICE OF NAVAL RESEARCH, ARLINGTON, VA. 22217 (11) 25 Jul 79
12. REPORT DATE: July 25, 1979
13. NUMBER OF PAGES: 12
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): (12) 15 p.
15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report): This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report): (14) TR-133
18. SUPPLEMENTARY NOTES:
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): Adaptive estimation; conditional reference sets; Monte Carlo swindles; robustness
20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

In adaptive estimation, it is often considered that an estimator has made a mistake if the component estimator chosen for use is not the most efficient for the distribution sampled. Theoretical and simulation results point to a fallacy in this line of thought. The Monte Carlo study involves extension of the Princeton Swindle to distributions conditional on a location- and scale free statistic, and to the uniform. The results give a partial explanation for the sometimes surprising robustness of adaptive L-estimators.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

404 198

alt